TASK -01

Create a bar chart or histogram to visualize the distribution of a categorical or continuous variable like age or gender distribution in a population.

1) import the Libraries

In [28]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

2) Download and Load Dataset

In [42]:
```
df = pd.read_csv("dataset1.csv", skiprows=4)
df
```

Out[42]:

	Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	1962	1963	1964	1
0	Aruba	ABW	Population, total	SP.POP.TOTL	5.492200e+04	5.557800e+04	5.632000e+04	5.700200e+04	5.761900e+04	5.819000e
1	Africa Eastern and Southern	AFE	Population, total	SP.POP.TOTL	1.300757e+08	1.335349e+08	1.371717e+08	1.409455e+08	1.449041e+08	1.490335e
2	Afghanistan	AFG	Population, total	SP.POP.TOTL	9.035043e+06	9.214083e+06	9.404406e+06	9.604487e+06	9.814318e+06	1.003601e
3	Africa Western and Central	AFW	Population, total	SP.POP.TOTL	9.763092e+07	9.970667e+07	1.018548e+08	1.040892e+08	1.063884e+08	1.087726e
4	Angola	AGO	Population, total	SP.POP.TOTL	5.231654e+06	5.301583e+06	5.354310e+06	5.408320e+06	5.464187e+06	5.521981e
5	Albania	ALB	Population, total	SP.POP.TOTL	1.608800e+06	1.659800e+06	1.711319e+06	1.762621e+06	1.814135e+06	1.864791e

In [ ]:
```
3) Clean the Data
```

In [41]:
```
population_df = df[['Country Name', '2022']].dropna()
population_df.columns = ['Country', 'Population']
population_df
```

Out[41]:

	Country	Population
0	Aruba	1.073100e+05
1	Africa Eastern and Southern	7.318214e+08
2	Afghanistan	4.057884e+07
3	Africa Western and Central	4.973872e+08
4	Angola	3.563503e+07
5	Albania	2.777689e+06
6	Andorra	7.970500e+04
7	Arab World	4.713521e+08
8	United Arab Emirates	1.007498e+07
9	Argentina	4.540790e+07
10	Armenia	2.969200e+06
11	American Samoa	4.834200e+04
12	Antigua and Barbuda	9.284000e+04
13	Australia	2.601440e+07
14	Austria	9.041851e+06
15	Azerbaijan	1.014176e+07
16	Burundi	1.332110e+07
17	Belgium	1.168021e+07
18	Benin	1.375950e+07
19	Burkina Faso	2.250904e+07
20	Bangladesh	1.693849e+08
21	Bulgaria	6.465097e+06
22	Bahrain	1.524693e+06
23	Bahamas, The	3.975380e+05
24	Bosnia and Herzegovina	3.204802e+06
25	Belarus	9.228071e+06

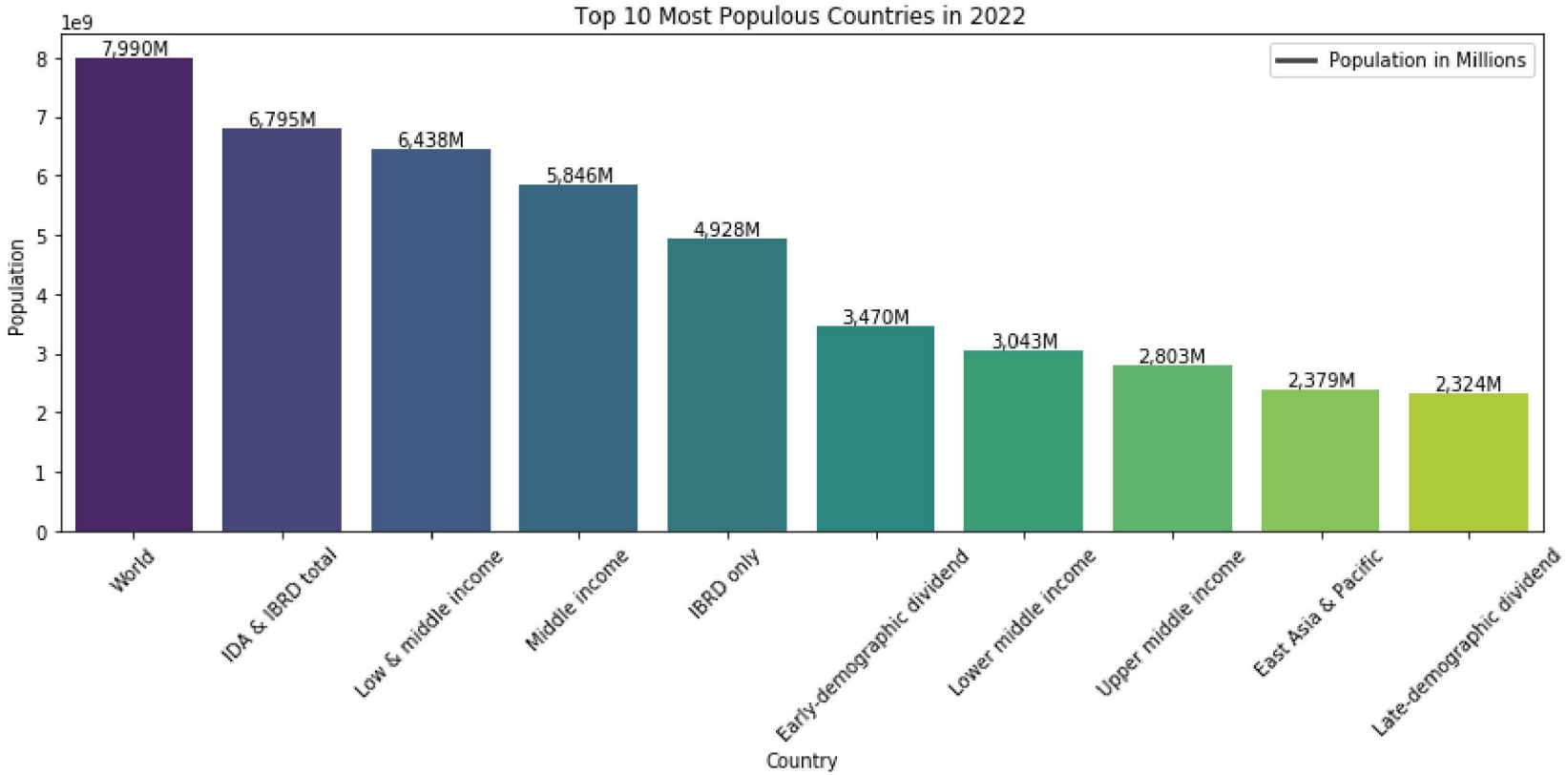
	Country	Population
26	Belize	4.027330e+05
27	Bermuda	6.474900e+04
28	Bolivia	1.207715e+07
29	Brazil	2.103064e+08
...	...	...
236	Latin America & the Caribbean (IDA & IBRD coun...	6.378378e+08
237	Timor-Leste	1.369295e+06
238	Middle East & North Africa (IDA & IBRD countries)	7.114650e+08
239	Tonga	1.050420e+05
240	South Asia (IDA & IBRD)	1.648010e+09
241	Sub-Saharan Africa (IDA & IBRD countries)	1.229209e+09
242	Trinidad and Tobago	1.365805e+06
243	Tunisia	1.211933e+07
244	Turkiye	8.497991e+07
245	Tuvalu	9.992000e+03
246	Tanzania	6.471182e+07
247	Uganda	4.731272e+07
248	Ukraine	4.104877e+07
249	Upper middle income	2.803432e+09
250	Uruguay	3.390913e+06
251	United States	3.340173e+08
252	Uzbekistan	3.493896e+07
253	St. Vincent and the Grenadines	1.020460e+05
254	Venezuela, RB	2.821302e+07
255	British Virgin Islands	3.831900e+04
256	Virgin Islands (U.S.)	1.054130e+05
257	Viet Nam	9.968066e+07

	Country	Population
258	Vanuatu	3.130460e+05
259	World	7.990400e+09
260	Samoa	2.152610e+05
261	Kosovo	1.768096e+06
262	Yemen, Rep.	3.822288e+07
263	South Africa	6.237841e+07
264	Zambia	2.015294e+07
265	Zimbabwe	1.606906e+07

265 rows × 2 columns

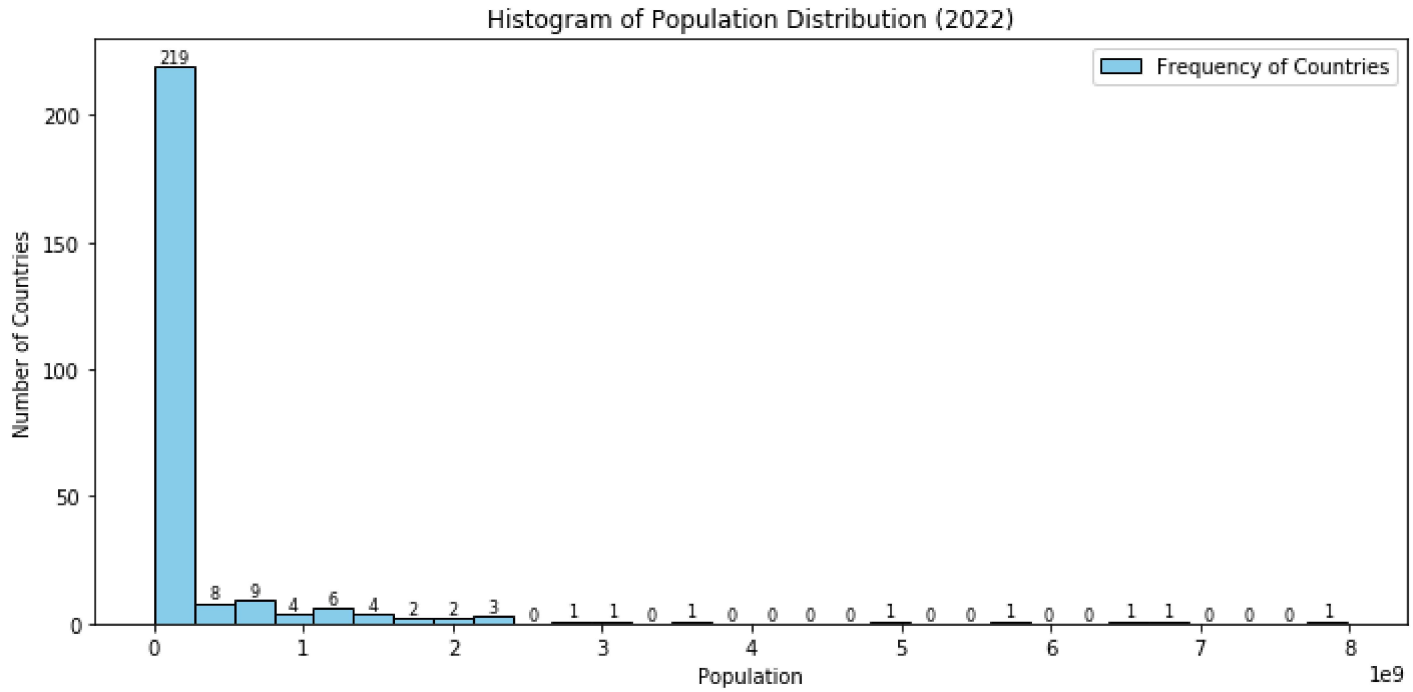
4) Bar Chart – Top 10 Countries

In [37]:
```
top10 = population_by_country.sort_values(by="Population", ascending=False).head(10)
plt.figure(figsize=(12, 6))
ax = sns.barplot(data=top10, x="Country", y="Population", palette="viridis")
plt.title("Top 10 Most Populous Countries in 2022")
plt.ylabel("Population")
plt.xlabel("Country")
plt.xticks(rotation=45)
for p in ax.patches:
    ax.annotate(f'{int(p.get_height()/1e6):,}M',(p.get_x() + p.get_width() / 2., p.get_height()),ha='center', v
plt.legend(["Population in Millions"], loc='upper right')
plt.tight_layout()
plt.show()
```



5) Histogram - All Countries Population

In [39]:
```
plt.figure(figsize=(10, 5))
n, bins, patches = plt.hist(population_by_country["Population"], bins=30, color="skyblue", edgecolor="black")
for i in range(len(patches)):
    plt.text(patches[i].get_x() + patches[i].get_width() / 2, patches[i].get_height(), str(int(patches[i].get_h
plt.title("Histogram of Population Distribution (2022)")
plt.xlabel("Population")
plt.ylabel("Number of Countries")
plt.legend(["Frequency of Countries"], loc='upper right')
plt.tight_layout()
plt.show()
```



In [ ]: